MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA

JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Text/Multimedia Mining and Semantic Technologies

Prof. Dr. Dunja Mladenić
Dr. Blaž Fortuna

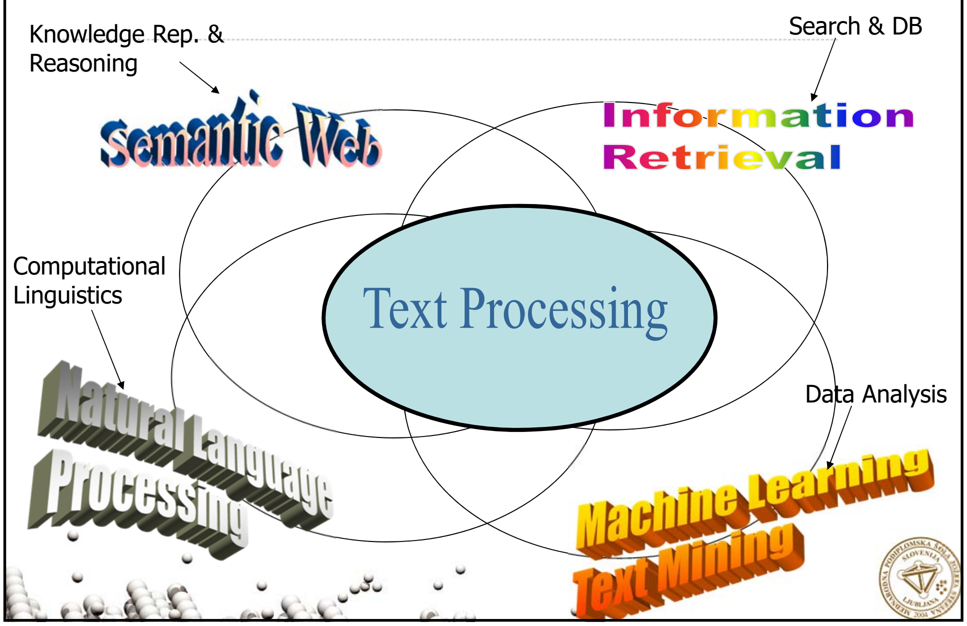
J. Stefan Institute
and
J. Stefan Postgraduate School

www.mps.si

Data Analytics

- Find interesting regularities in (large, text/multimedia) data
  – interesting: non-trivial, hidden, previously unknown and potentially useful
- Data modeling for prediction and/or description
- Data visualization

Related research areas



Creativity in Research

Follow the general process of creation:

- intention initiates the process
- consciousness brings focus and put it in broader context
- intelligence to ensure conditions
- energy to manifest

Applied on research:

- spark of idea
- seeing possible consequences
- articulation/analysis
- implementation

Approaching research problem

- Idea/intention – intuitive spark
  - See the idea in a broader context
- Strategy – intellectual analysis
  - Understand the practicalities
    - domain, available data, state-of-the-art methods
  - Identify the needed steps
    - including resources (knowledge, equipment, time, …)
- Implementation – practical action
  - Develop an approach/theory
  - Evaluate and revise as needed
  - Reflect on the lessons learned

Approaching … lem

- Idea/intention – intuitive s…
  - See the idea in a broader c…
- Strategy – intellectual a…
  - Understand the practicalities
    - domain, available data, state-…
  - Identify the needed steps
    - including resources (knowledge, equipment, time, …)
- Implementation – pr…
  - Develop an approach…
  - Evaluate and revise as needed
  - Reflect on the lessons learn…

improve effectiveness & wellbeing of users

mobile device gives personalized suggestions for activity

target leaders using demographics, position, skills, daily activities

knowledge base of activities, user feedback,…

unsupervised learning of user profile, semantically annotated activities, active learning for suggestions

questionnaire for user validation, visualization of profiles

Tasks to address

- Issue to address determines task
  - Search for information (text, image, video clip, audio)
  - Learning model (un-, semi-, supervised)
  - Summarization
  - Translation
  - Visualization
  - …

Representation to use

Available data and task influence representation

- text can be represented on different level of granularity
  - character-level, word level,… to logic
- natural languages of texts
  - mono-lingual, multi-lingual, cross-lingual
- text combined with other data
  - multimodal data representation

Techniques to apply

Task and practical requirements influences techniques

- from manual work to machine learning and reasoning
- trade-off between scalability (data storage) and latency (processing speed)
- to consider quality, resources, standards, …

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Representing Text Data

www.mps.si

Levels of text representations

- Character (character n-grams and sequences)
- Words (stop-words, stemming, lemmatization)
- Phrases (word n-grams, proximity features)
- Part-of-speech tags
- Taxonomies / thesauri
- Vector-space model

Lexical

- Language models
- Full-parsing
- Cross-modality

Syntactic

- Collaborative tagging / Web2.0
- Learning Features – word embedding
- Templates / Frames
- Ontologies / First order theories

Semantic

Levels of text representations

- Character (character n-grams and sequences)
- Words (stop-words, stemming, lemmatization)
- Phrases (word n-grams, proximity features)
- Part-of-speech tags
- Taxonomies / thesauri
- Vector-space model

Lexical

- Language models
- Full-parsing
- Cross-modality

Syntactic

- Collaborative tagging / Web2.0
- Learning Features – word embedding
- Templates / Frames
- Ontologies / First order theories

Semantic

Language identification, Copy detection

Named entity extraction (names of people, places, organizations)

Text categorization, Clustering, Search , Summarization, …

Spam filtering, Machine translation

Multilingual search, Associating text with images, …

Unifying semantics of data

Reasoning, Semantic search

Word level

- The most common representation of text used for many techniques
  - …there are many tokenization software packages which split text into the words
- Important to know:
  - Word is well defined unit in western languages – e.g. Chinese has different notion of semantic unit

Stop-words

- Stop-words are words that from non-linguistic view do not carry information
  - …they have mainly functional role
  - …usually we remove them to help the methods to perform better
- Stop words are language dependent – examples:
  - **English**: A, ABOUT, ABOVE, ACROSS, AFTER, AGAIN, AGAINST, ALL, ALMOST, ALONE, ALONG, ALREADY, ...
  - **Dutch**: de, en, van, ik, te, dat, die, in, een, hij, het, niet, zijn, is, was, op, aan, met, als, voor, had, er, maar, om, hem, dan, zou, of, wat, mijn, men, dit, zo, ...
  - **Slovenian**: A, AH, AHA, ALI, AMPAK, BAJE, BODISI, BOJDA, BRŽKONE, BRŽČAS, BREZ, CELO, DA, DO, ...

Stemming and lemmatization

- Different forms of the same word are usually problematic for text data analysis, because they have different spelling and similar meaning (e.g. learns, learned, learning,…)
- Stemming is a process of transforming a word into its stem
  - (universe, university, universities, university's, universal) –> univers
- Lemmatization transforms word into its normalized form
  - universe -> universe, (university, universities, university's) -> university, universal –> universal
- …stemming provides an inexpensive mechanism to merge words with similar meaning

Stemming

- For English is mostly used Porter stemmer at http://www.tartarus.org/~martin/PorterStemmer/
- Example cascade rules used in English Porter stemmer

– ATIONAL	-> ATE	relational -> relate
– TIONAL	-> TION	conditional -> condition
– ENCI	-> ENCE	valenci -> valence
– ANCI	-> ANCE	hesitanci -> hesitance
– IZER	-> IZE	digitizer -> digitize
– ABLI	-> ABLE	conformabli -> conformable
– ALLI	-> AL	radicalli -> radical
– ENTLI	-> ENT	differentli -> different
– ELI	-> E	vileli -> vile
– OUSLI	-> OUS	analogousli -> analogous

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA

JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Example tasks

www.mps.si

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA

JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Communication Analysis

GROBELNIK, Marko, MLADENIĆ, Dunja, FORTUNA, Blaž. Semantic technology for capturing communication inside an organisation. *IEEE internet computing*, 2009, vol. 13, no. 4, pp. 59-66.

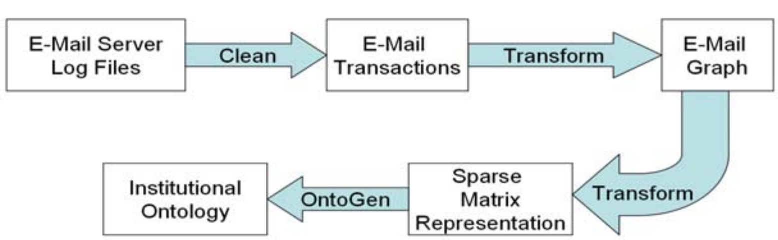
www.mps.si

Social networks

- Social networks can be also potential source of data for machine learning and building semantic structures
  - …conceptually they share similar underlying structure as text – namely, the underlying distribution is generated by power-law
- In the next slides we show how social networks can be modeled using unsupervised techniques

Analysis of e-mail graph

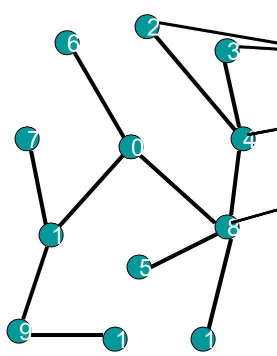
- An e-mail graph can be analyzed in the following 5 major steps:
  1. Starting with log files from an e-mail server where the data include information about e-mail transactions with the fields: **sender** and the **list of receivers**.
  2. After cleaning we get the data in the form of **e-mail transactions** which include e-mail addresses of **sender** and **receiver**.
  3. From a set of **e-mail transactions** we construct a **graph** where vertices are e-mail addresses connected if there is a transaction between them
  4. **E-mail graph** is transformed into a **sparse matrix** allowing to perform data manipulation and analysis operations
  5. **Sparse matrix** representation of the graph is analyzed with **ontology learning** tools producing an **ontological structure** corresponding to the **organizational structure** of the institution where e-mails came from.



Graph transformation into a set of sparse matrix

- Graph with N *vertices is transformed into N*N sparse matrix where:*
  - *…Xth row represents information for Xth vertex*
  - *…Xth row has nonzero components for:*
    - *Xth vertex itself and*
    - *Xth vertex's neighbors on the distance D (e.g. 1, 2, 3)*
  - *Intuitively, Xth row represents numerically "neighborhood" of the Xth vertex within the graph:*
    - *Xth element in the Xth row has weight 1*
    - *…elements representing neighbors have lower weights relative to the distance (d) from the Xth vertex (1/(2^d))*
      - *(e.g. 1, 0.5, 0.25, 0.125, …)*

Graph transformation into sparse matrix (example)



	0	1	2	3	4	5	6	7	8	9	10	11
0	1	0.5			0.25	0.25	0.5	0.25	0.5	0.25		0.25
1	0.5	1					0.25	0.5	0.25	0.5	0.25	
2			1	0.25	0.5				0.25			
3			0.25	1	0.5				0.25			
4	0.25		0.5	0.5	1	0.25			0.5			0.25
5	0.25				0.25	1			0.5			0.25
6	0.5	0.25					1		0.25			
7	0.25	0.5						1		0.25		
8	0.5	0.25	0.25	0.25	0.5	0.5	0.25		1			0.5
9	0.25	0.5						0.25		1	0.5	
10		0.25								0.5	1	
11	0.25				0.25	0.25			0.5			1

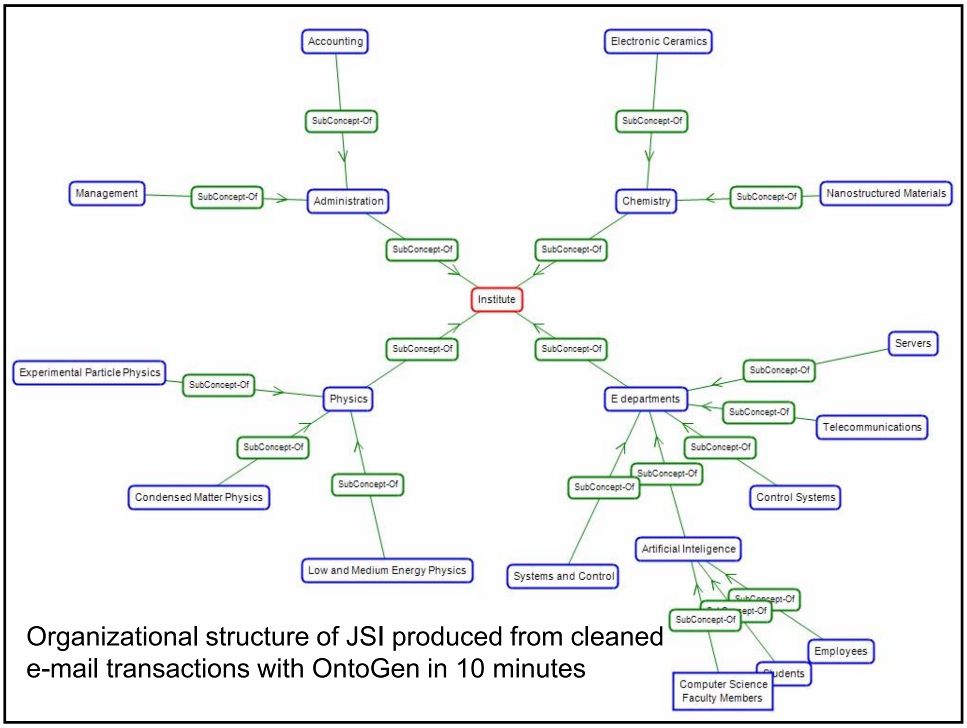
Transforming Graph into Matrix

Data used for experimentation

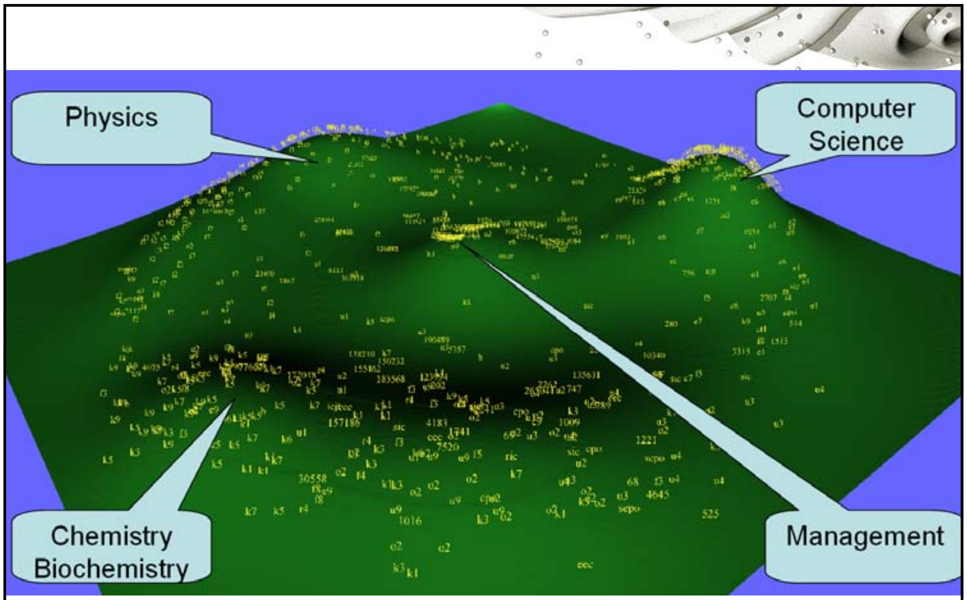
- The data is the collection of log files with e-mail transactions from local e-mail spam filter software Amavis (http://www.amavis.org/):
  - Each line of the log files denotes one event at the spam filter software
  - We were interested in the events on successful e-mail transactions
    - ...having information on **time**, **sender**, and **list of receivers**
  - An example of successful e-mail transaction is the following line:
    - `2005 Mar 28 13:59:05 patsy amavis[33972]: (33972-01-3) Passed CLEAN, [217.32.164.151] [193.113.30.29] <john.nj.davies@bt.com> -> <marko.grobelnik@ijs.si>, Message-ID: <21DA6754A9238B48B92F39637EF307FD0D4781C8@i2km41-ukdy.domain1.systemhost.net>, Hits: -1.668, 6389 ms`

Some statistics about the data

- The log files include e-mails for 19 months:
  - …this sums up to 12.8Gb of data.
  - After filtering out successful e-mail transactions it remains 564Mb
    - …which contains approx. 2.7 million of successful e-mail transitions used for further processing
  - The whole dataset contains references to approx. 45000 e-mail addresses
    - …after the data cleaning phase the number is reduced to approx. 17000 e-mail addresses
    - …out of which 770 e-mail addresses are internal from the home institution (with local domain name)



Organizational structure of JSI produced from cleaned e-mail transactions with OntoGen in 10 minutes



Organizational structure of JSI visualized from e-mail transactions with Document-Atlas

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA

JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Cross-lingual Event Extraction

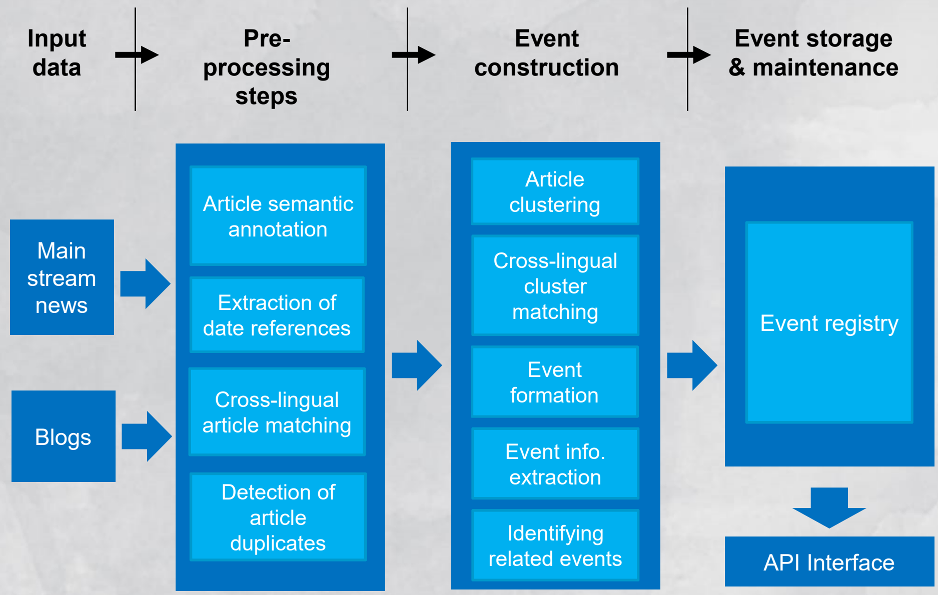
Leban, G., Fortuna, B., Brank, J., & Grobelnik, M. Event Registry – learning about world events from news, In Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, 2014.

www.mps.si

Cross-lingual event extraction from news

- News articles in different languages
- Preprocessed and annotated enabling cross-lingual article matching
- Event extraction by cross-lingual clustering
- Enable advanced search and rich visualization options

Event extraction pipeline



Event formation from text stream

- Event is formed from one or more linked clusters
  - as clusters evolve, they can be added or removed from the event
- Each event is assigned a unique id
- Extract event information using the articles
  - to answer questions what, when, where, who
  - title and the 1st paragraph of the medoid article
  - Date - the most frequent or average article date

Event Registry

Event Registry http://eventregistry.org/

- Database of all detected events + extracted information about them
- Provides API to search for events

- Event data is also provided in structured form
  - Use of BBC Storyline ontology
- SPARQL endpoint:
  - http://eventregistry.org/rdf/search

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA

JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Techniques for Data Modeling

www.mps.si

Basic approaches to modeling using machine learning methods

- **Supervised learning** (classification)
- **Semi-supervised learning** (transduction, active learning)
- **Unsupervised learning** (clustering)

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA

JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

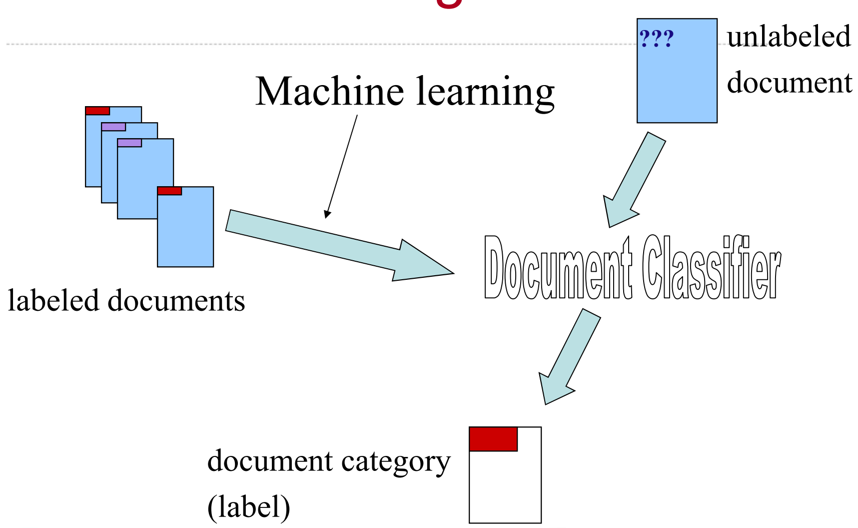
Supervised Learning

www.mps.si

Document Categorization Task

- **Given:** set of documents labeled with content categories
- **The goal:** to build a model which would automatically assign right content categories to new unlabeled documents.
- Content categories can be:
  - unstructured (e.g., Reuters) **or**
  - structured (e.g., Yahoo, DMoz, Medline)

Document categorization



Measuring success - Model quality estimation

The truth, and

..the whole truth

$$Precision(M,targetC) = P(targetC \mid \overline{targetC})$$

$$Recall(M,targetC) = P(\overline{targetC} \mid targetC)$$

$$Accuracy(M) = \sum_i P(\overline{C_i}) \times Precision(M,C_i)$$

$$F_\beta(M,targetC) = \frac{(1+\beta^2)Precision(M,targetC) \times Recall(M,targetC)}{\beta^2 Precision(M,targetC) + Recall(M,targetC)}$$

- Classification accuracy
- Break-even point (precision=recall)
- F-measure (precision, recall = sensitivity)

Algorithms for learning document classifiers

- Popular algorithms for text categorization:
  - Support Vector Machines
  - Logistic Regression
  - Perceptron algorithm
  - Naive Bayesian classifier
  - Winnow algorithm
  - Nearest Neighbour
  - ....
- Unlike decision tree and rule learning algorithms, these are mainly non-symbolic learning algorithms

Example learning algorithm: Perceptron

Input:
- set of documents ***D*** in the form of (e.g. TFIDF) numeric vectors
- each document has label +1 (positive class) or -1 (negative class)

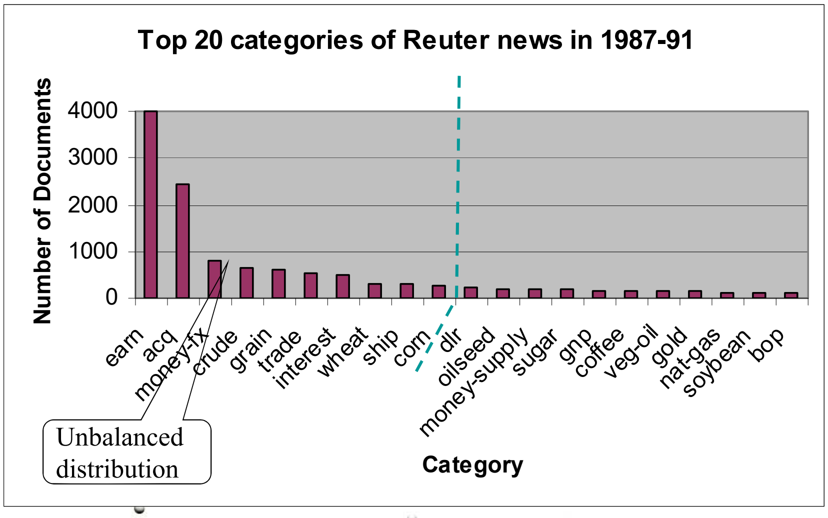
Output:
- linear model w_i (one weight per word from the vocabulary)

Algorithm:
- **Initialize** the model w_i by setting word weights to 0
- **Iterate** through documents N times
  - **For** document ***d*** from ***D***
    - // Using current model w_i classify the document ***d***
    - **if** sum(d_i *w_i) >= 0 **then** classify document as positive
    - **else** classify document as negative
    - **if** document classification is wrong **then**
      - *// adjust weights of all words occurring in the document*
      - $w_{t+1} = w_t$ +sign(true-class) * Beta (input parameter Beta>0)
      - *// where sign(positive) = 1 and sign(negative) =-1*

Categorization to flat categories

Example data set used in research:
- Documents are classified by editors into one or more categories
- Publicly available set of Reuter news mainly from 1987:
  - 120 categories giving the document content, such as: earn, *acquire, corn, rice, jobs, oilseeds, gold, coffee, housing, income,...*
- Larger dataset available for research from 2000 having 830,000 Reuters news documents

Distribution of documents
(Reuters-21578)



Example of Perceptron model for Reuters category "Acquisition"

Feature	Positive Class Weight
STAKE	11.5
MERGER	9.5
TAKEOVER	9
ACQUIRE	9
ACQUIRED	8
COMPLETES	7.5
OWNERSHIP	7.5
SALE	7.5
OWNERSHIP	7.5
BUYOUT	7
ACQUISITION	6.5
UNDISCLOSED	6.5
BUYS	6.5
ASSETS	6
BID	6
BP	6
DIVISION	5.5

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA

JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Semi-supervised Learning

www.mps.si

Semi-supervised learning

Similar to supervised learning except that

- we have examples and only some of them are labeled
- we may have a human available for a limited time to provide labels of examples
  - …this corresponds to the situation where all the cartoons in our collection have descriptions, but only a few have label
  - …and occasionally we have a human for a limited time to respond the questions about the cartoons

Document categorization with only few labeled documents

- we have many documents but only some of them are labeled
- we may have a human available for a limited time to provide labels of documents

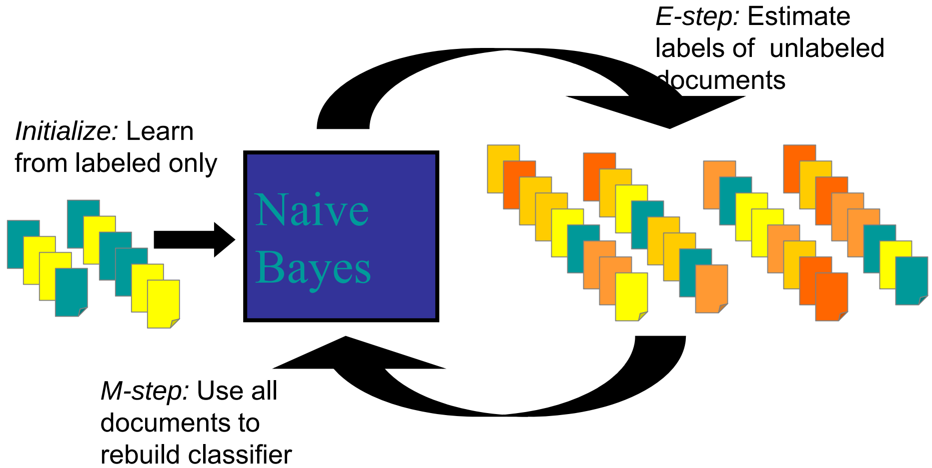
Approaches:

- Using unlabeled data
- Co-training
- Active learning

Using unlabeled data [Nigam et al., 2000]

- Given: a small number of labeled examples and a large pool of unlabeled examples, no human available
  - e.g., classifying news article as interesting or not interesting
- Approach description (EM + Naive Bayes):
  - train a classifier with only labeled documents,
  - assign probabilistically-weighted class labels to unlabeled documents,
  - train a new classifier using all the documents
  - iterate until the classifier remains unchanged

Using Unlabeled Data with Expectation-Maximization (EM)



- Guarantees local maximum a posteriori parameters

Co-training [Blum & Mitchell, 1998]

Theory behind co-training

- Possible to learn from unlabeled examples
- Value of unlabeled data depends on
  - How (conditionally) independent are the two representations of the same data
    - The more the better
  - The number of redundant inputs (features)
    - Expected error decreases exponentially with this number
- Disagreement on unlabeled data predicts true error

Better performance on labelling unlabeled data compared to EM approach

Bootstrap Learning to Classify Web Pages

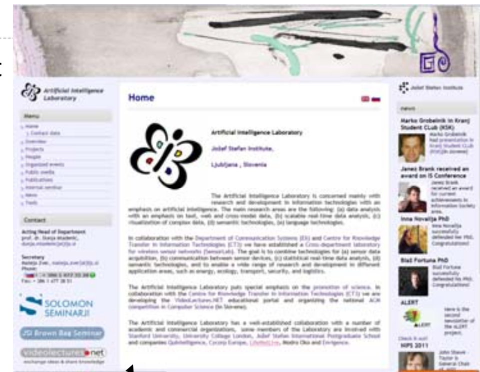
Document

Given: set of documents where each document is described by two independent sets of features (e.g. document text + hyperlinks anchor text)

few labeled and many unlabeled

Page Classifier

Link Classifier



Hyperlink to the document

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA

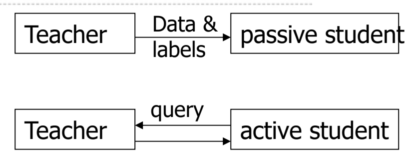
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Active Learning

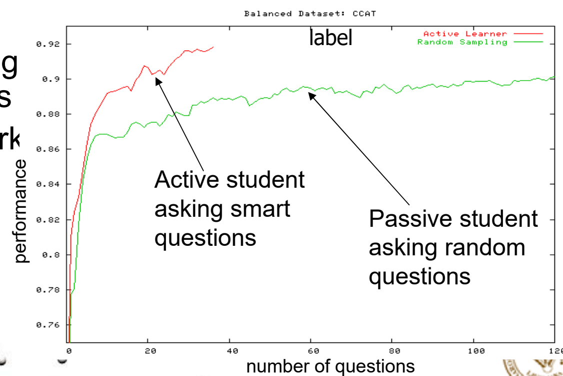
www.mps.si

Active Learning

- We use this methods whenever hand-labeled data are rare or expensive to obtain
- Interactive method



- Requests only labeling of “interesting” objects
- Much less human work needed for the same result compared to arbitrary labeling examples



Some approaches to Active Learning

- **Uncertainty sampling** (efficient)
  - select example closest to the decision hyperplane (or the one with classification probability closest to P=0.5) [Tong & Koller 2000]
- **Maximum margin ratio change**
  - select example with the largest predicted impact on the margin size if selected [Tong & Koller 2000]
- **Monte Carlo Estimation of Error Reduction**
  - select example that reinforces our current beliefs [Roy & McCallum 2001]
- **Random sampling** as baseline

- Experimental evaluation (using F1-measure) of the four listed approaches shown on three categories from Reuters-2000 dataset [Novak & Mladenic & Grobelnik, 2006]
  - average over 10 random samples of 5000 training (out of 500k) and 10k testing (out of 300k)examples
  - two of the methods a rather time consuming, thus we run them for including the first 50 unlabeled examples
  - experiments show that active learning is especially useful for unbalanced data

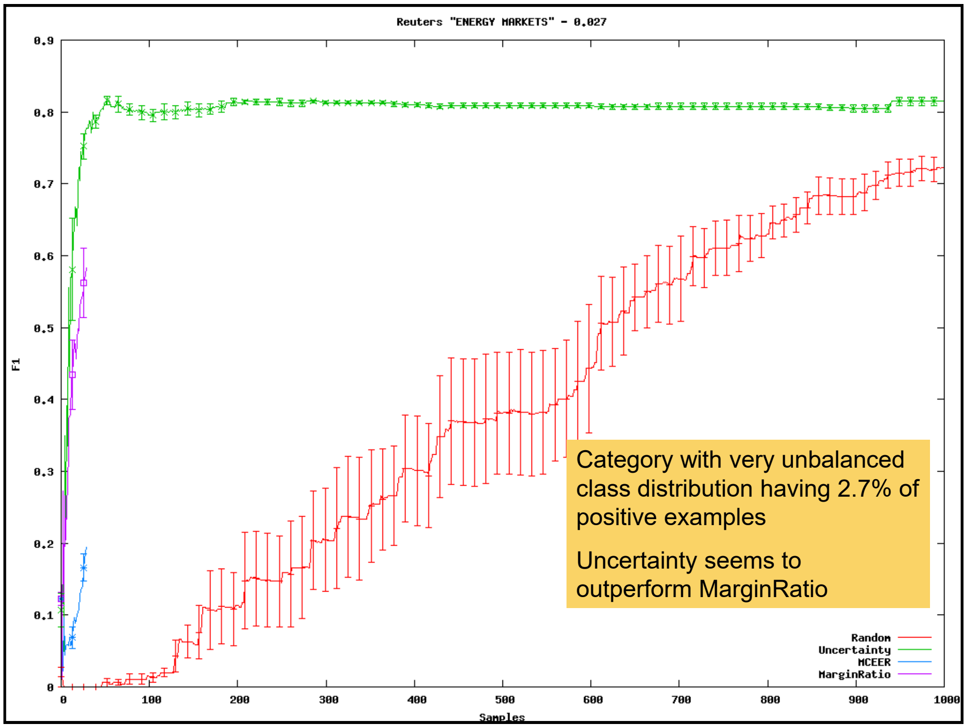
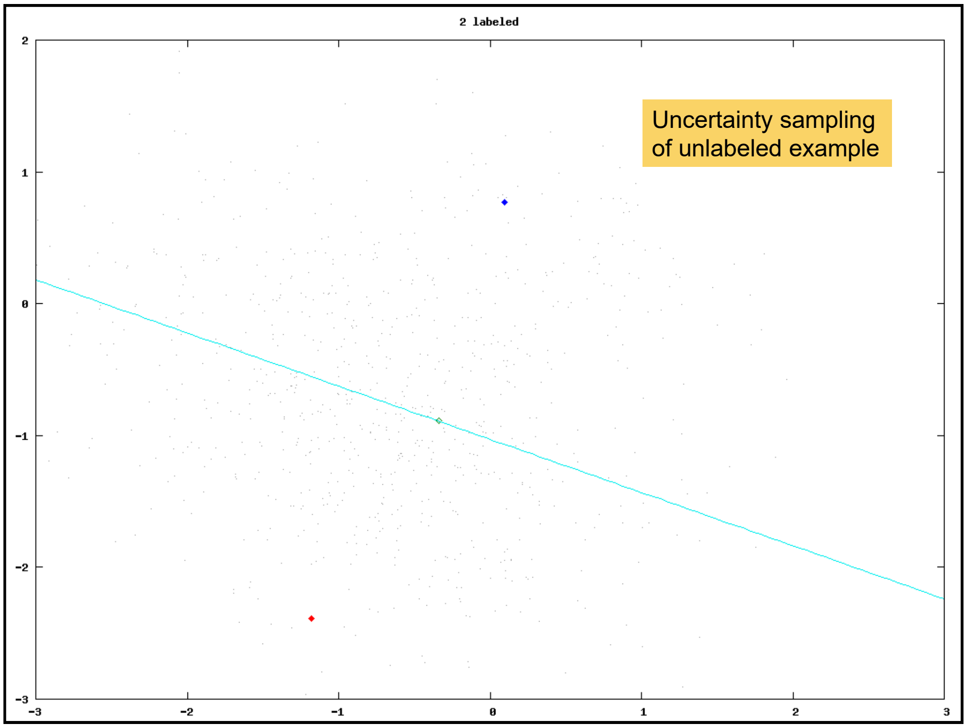


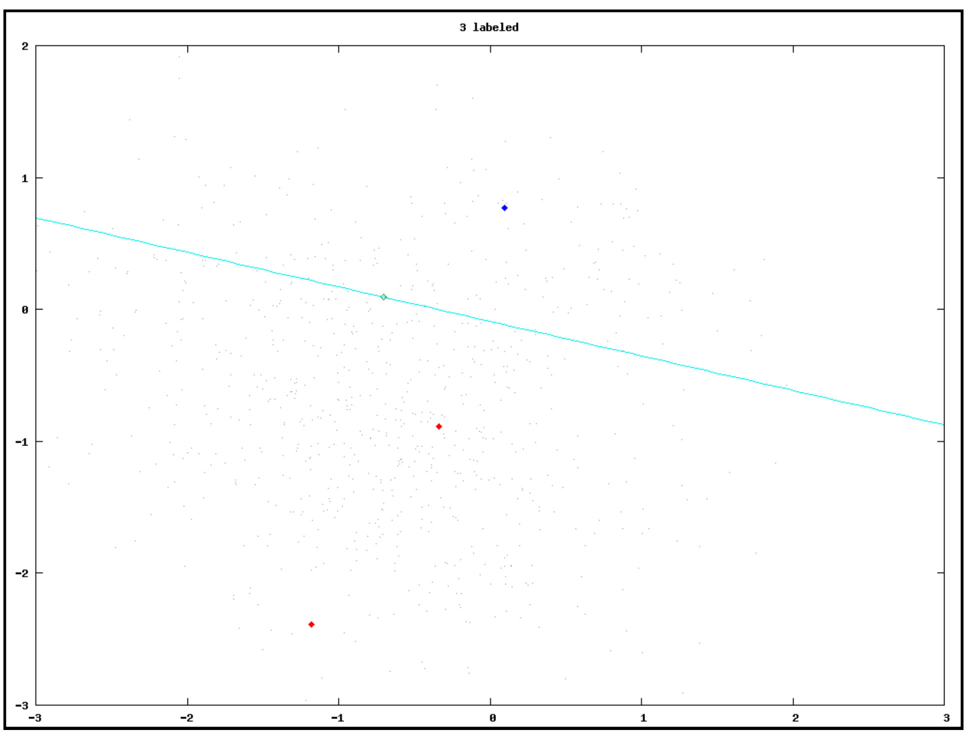
Illustration of Active learning

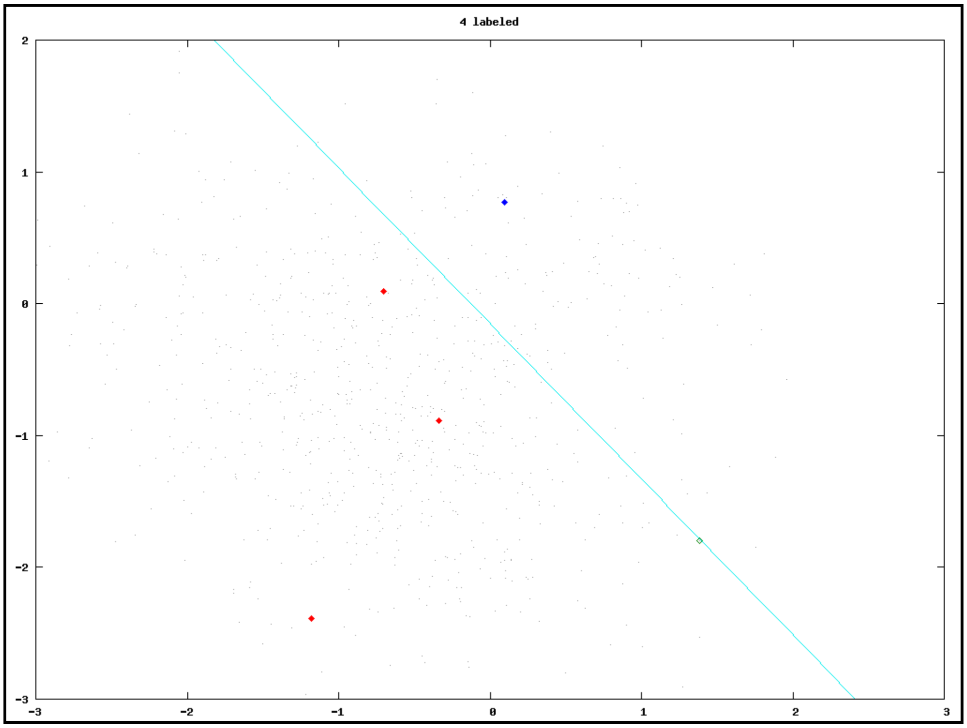
- starting with one labeled example from each class (red and blue)
- select one example for labeling (green circle)
- request label and add re-generate the model using the extended labeled data

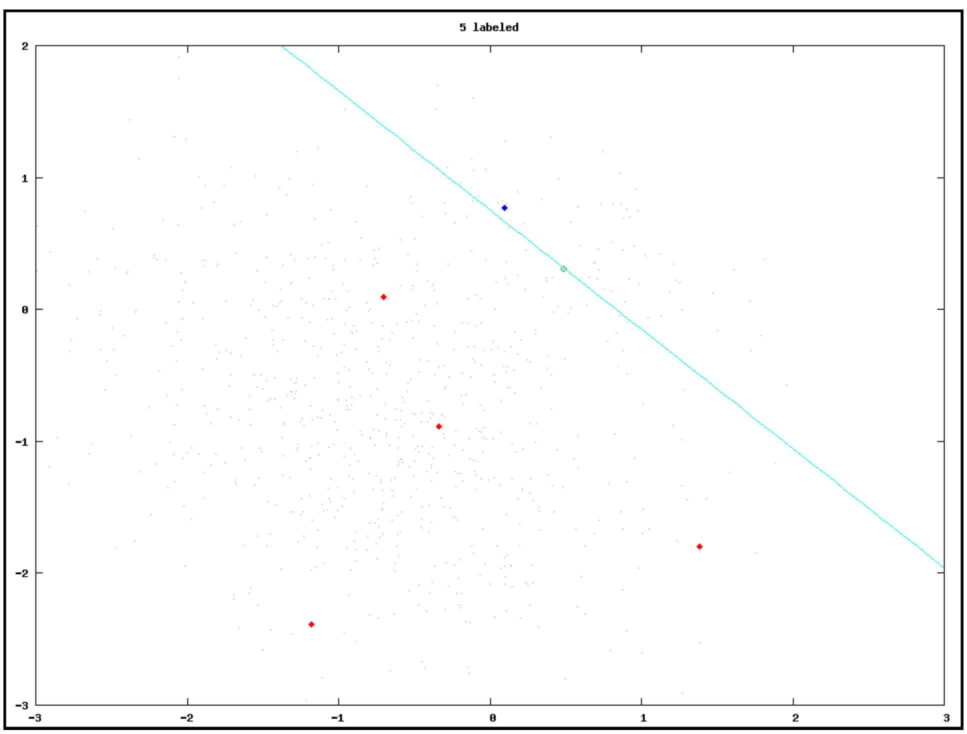
Illustration of linear SVM model using

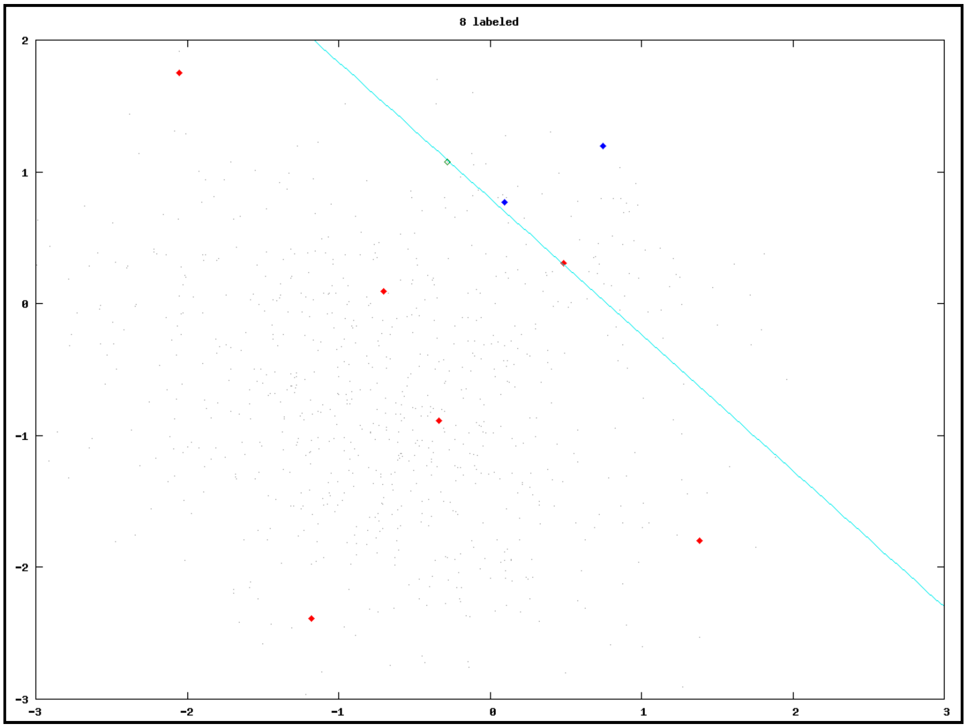
- arbitrary selection of unlabeled examples (random)
- active learning selecting the most uncertain examples (closest to the decision hyperplane)

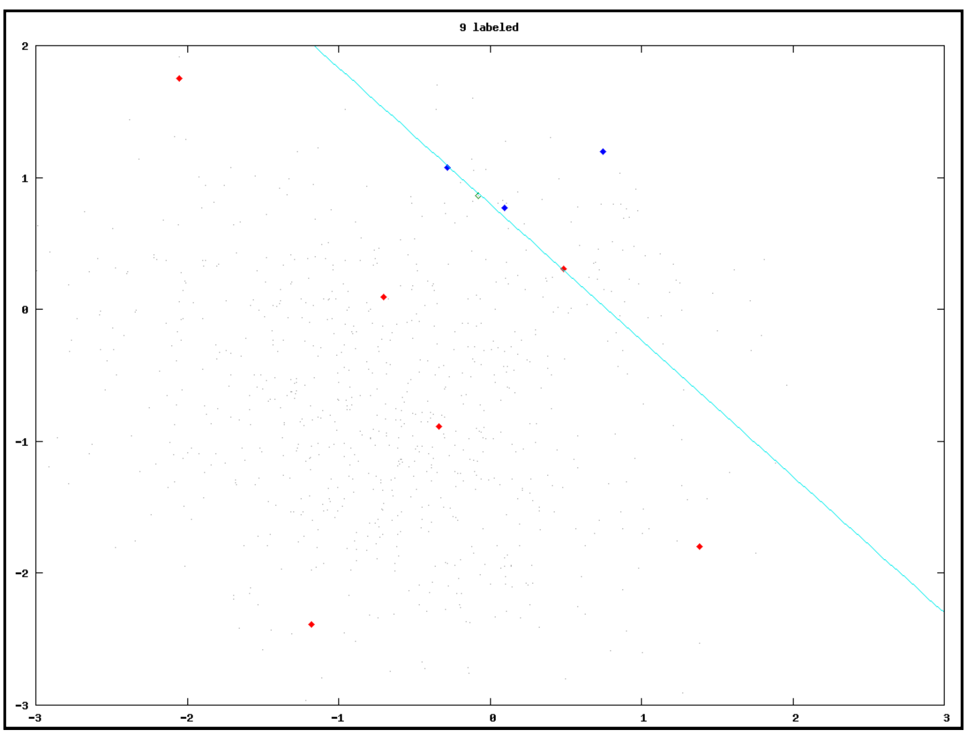

2 labeled
Uncertainty sampling
of unlabeled example

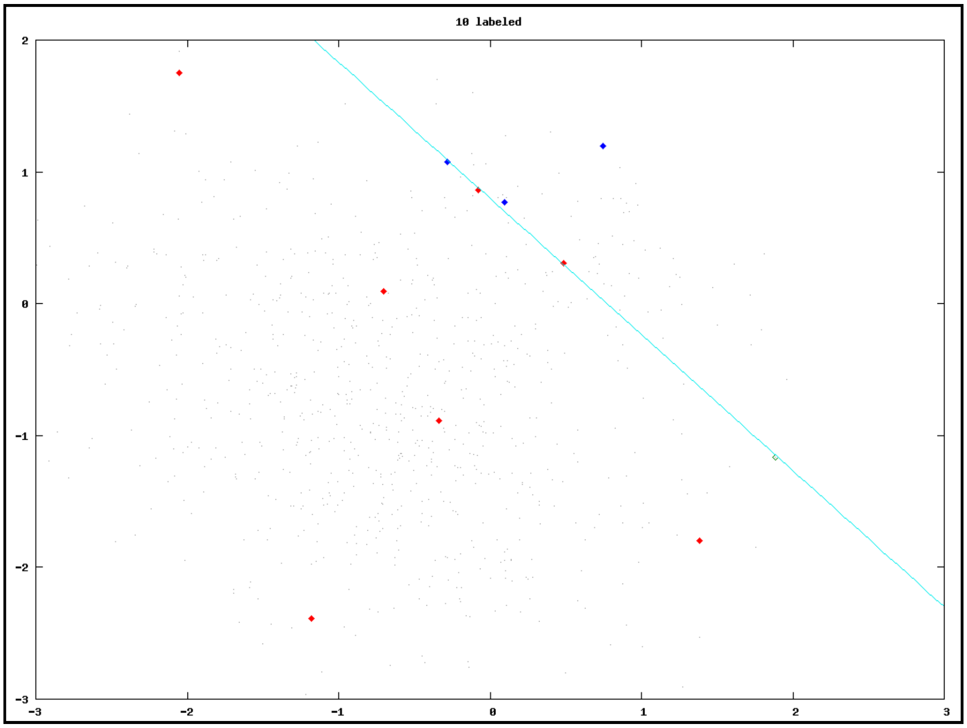

3 labeled

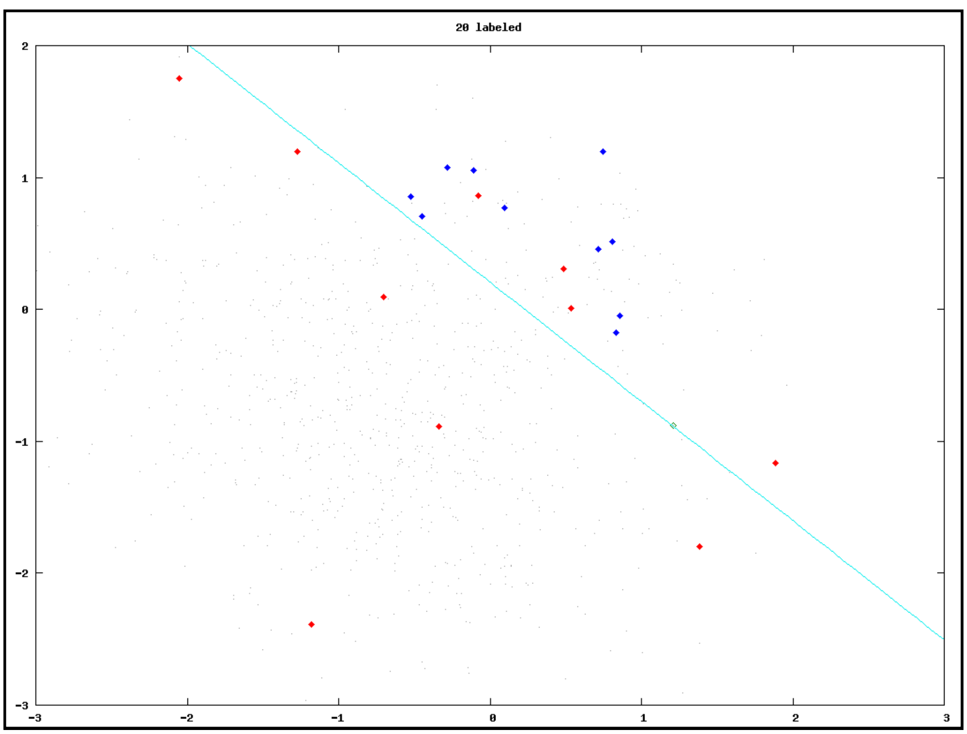

4 labeled
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

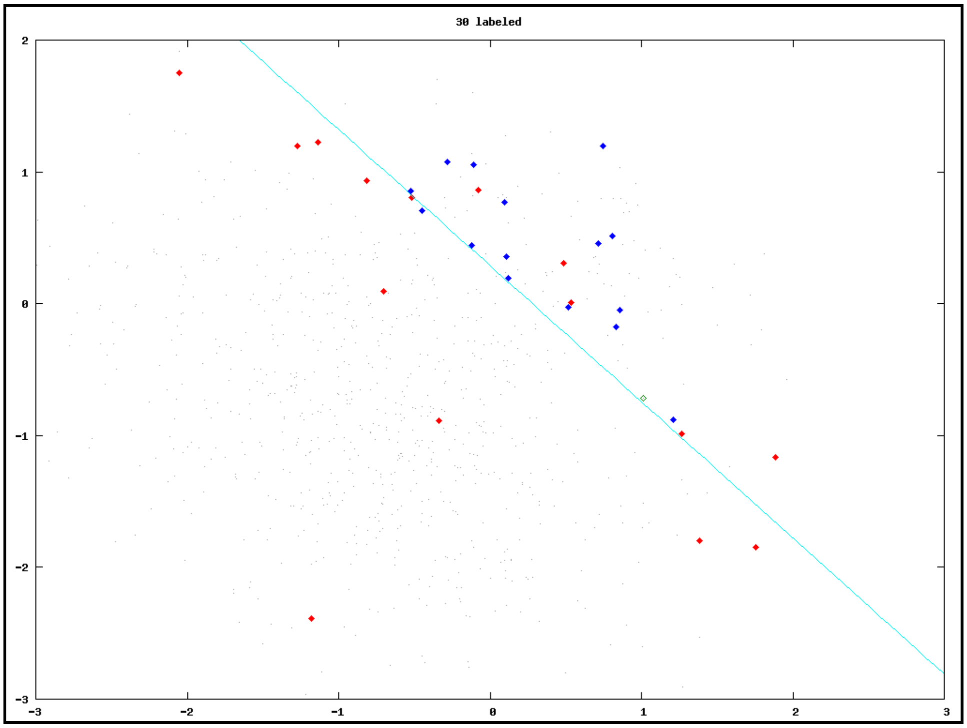

5 labeled
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

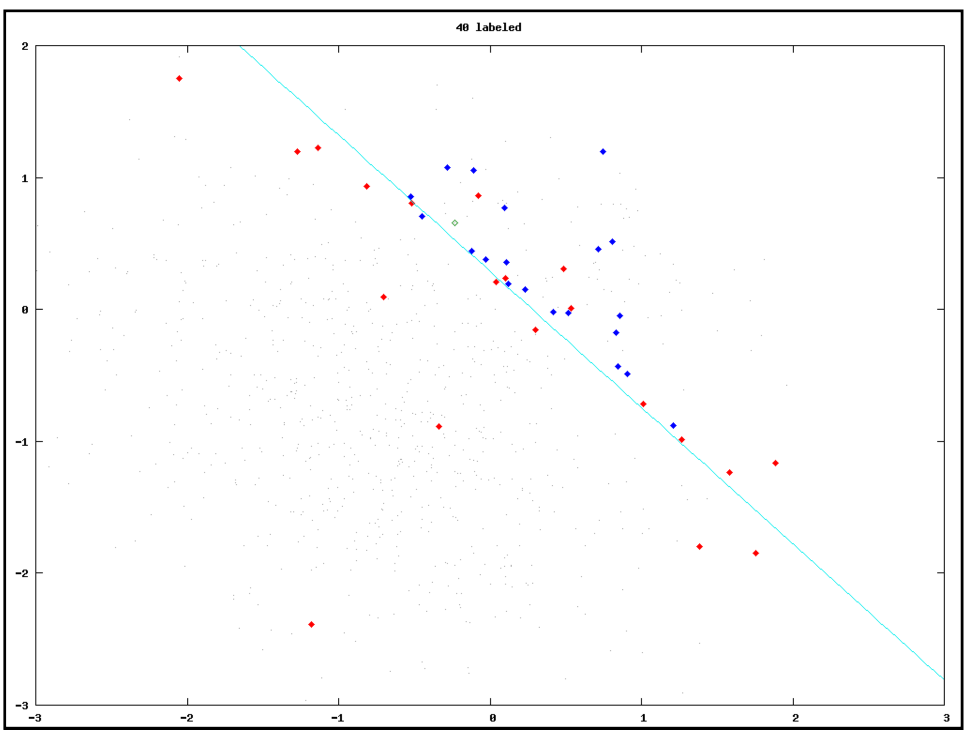

8 labeled
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

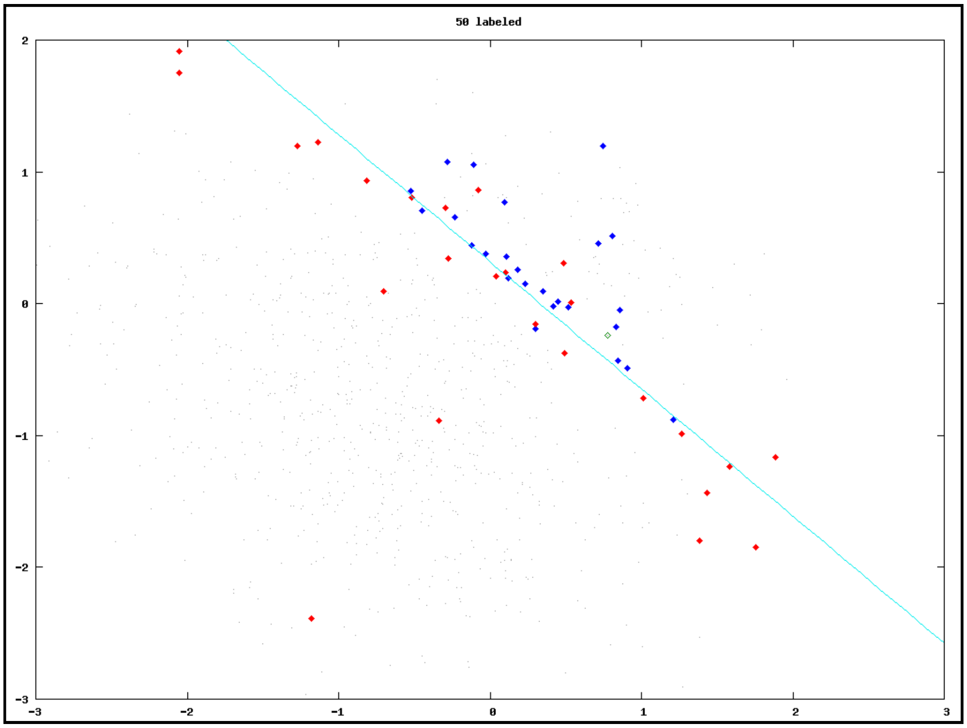

9 labeled
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

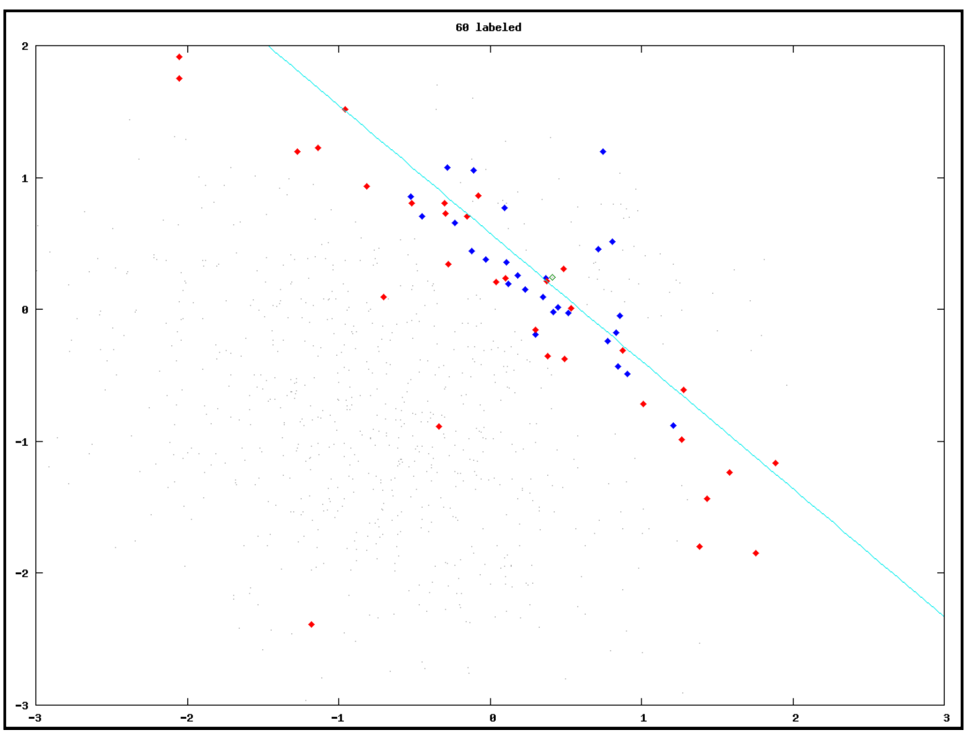

10 labeled
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

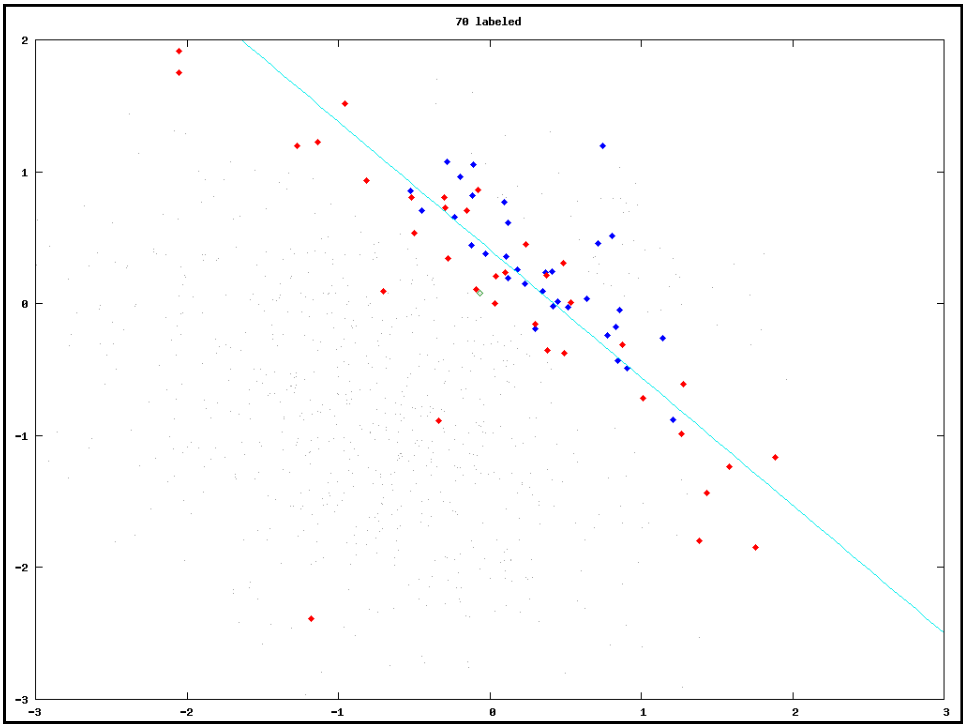

20 labeled
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

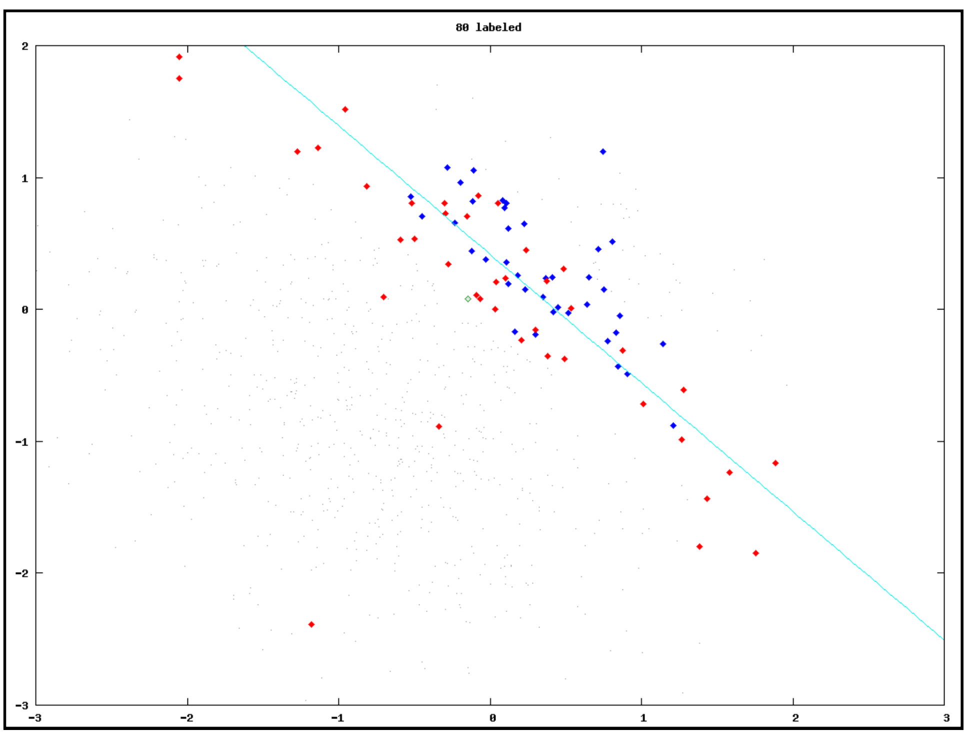

30 labeled
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

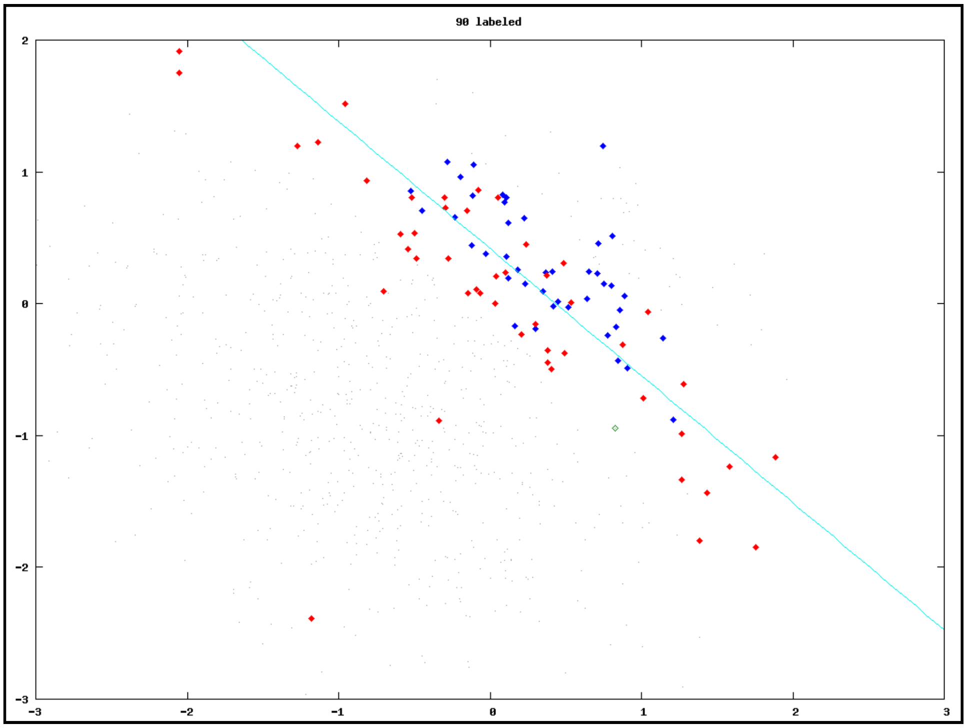

40 labeled
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

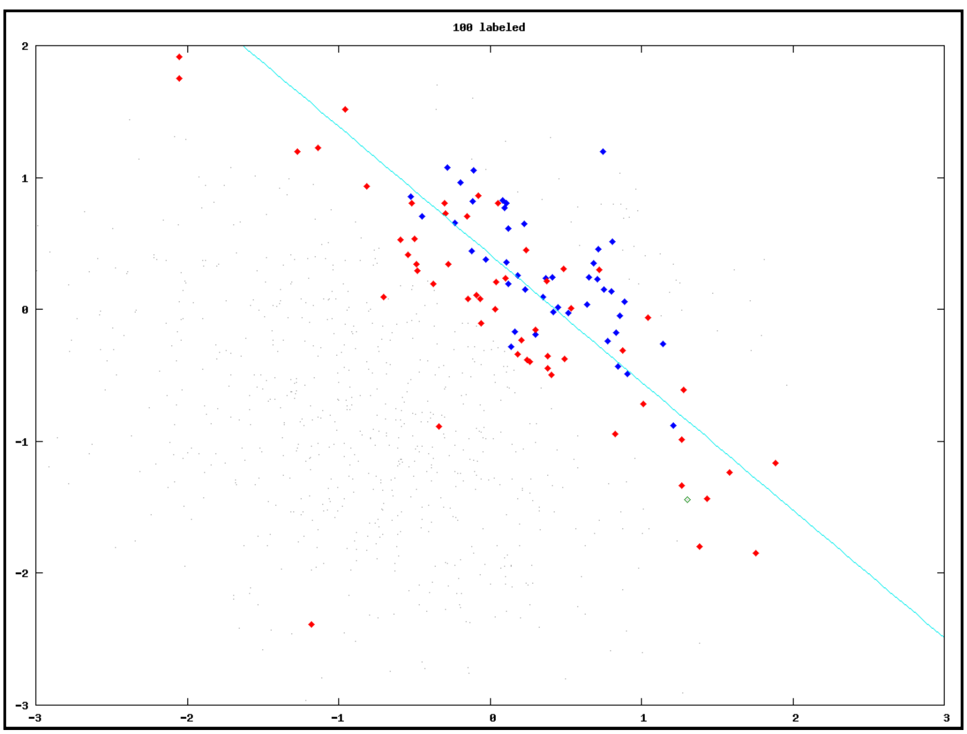

50 labeled
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

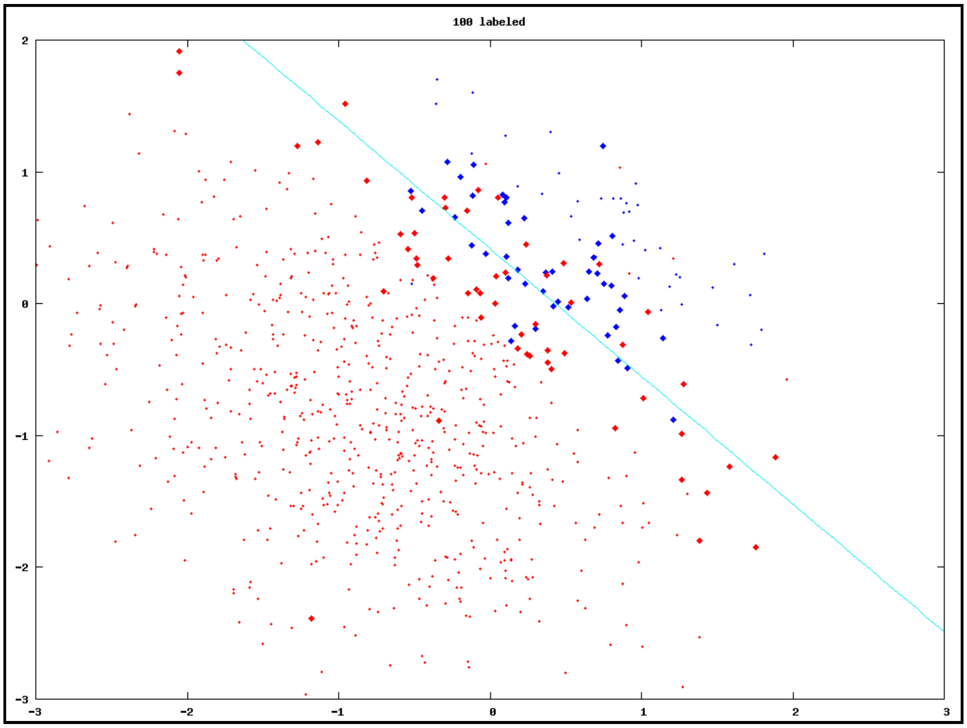

60 labeled
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3


70 labeled
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3


80 labeled
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3


90 labeled
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3


100 labeled
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3



Unsupervised Learning

Unsupervised learning

Document Clustering:

- Given is a set of documents
- The goal is: to cluster the documents into several groups based on some similarity measure
  - documents inside the group should be similar while documents between the groups should be different

Similarity measure plays a crucial role in clustering, on documents we use cosine similarity:

$$Cos(d_1, d_2) = \frac{d_1 \cdot d_2}{\|d_1\| \|d_2\|} = \frac{\sum_i x_{1i} x_{2i}}{\sqrt{\sum_j x_j^2} \sqrt{\sum_k x_k^2}}$$

Clustering methods

- Hierarchical
  - agglomerative – at each step merge two or more groups
  - divisive – at each step break the selected group into two or more groups
- Non hierarchical
  - requires specification of the number of clusters
  - optimization of the initial clustering (e.g., maximize similarity of examples inside the same group)
- Geometrical
  - map multidimensional space into two- or three-dimensional (e.g., principal component analysis)
- Graph-theoretical

K-Means clustering algorithm

- **Given**:
  - set of examples (e.g., TFIDF vectors of documents),
  - distance measure (e.g., cosine)
  - ***K*** (number of groups)
- **For each** of ***K*** groups initialize its centroid with a random document
- **While** not converging
  - Each document is assigned to the nearest group (represented by its centroid)
  - For each group calculate new centroid (group mass point, average document in the group)

Example of k-means clustering

	A	B	C	D	E
w1	1	1	1	0	0
w2	0	0	0	0	1
w3	1	0	1	0	0
w4	0	0	0	1	1
w5	1	1	0	0	0

K=2

1. Randomly select two examples, e.g., A, D to be representatives of two clusters I: A, II: D
2. Calculate similarity of other examples to the them B,I= 0.82, B,II= 0, C,I= 0.82, C,II= 0, E,I= 0, E,II= 0.7
3. Assign examples to the most similar cluster I: (A,B,C) II: (D,E)
4. Calculate the cluster centroid I: 1,0,0.67,0,0.67 II: 0,0.5,0,1,0
5. Calculate similarity of all the examples to the centroids A,I= 0.88, A,II= 0, B,I= 0.77, B,II= 0, C,I= 0.77, C,II= 0, D,I= 0, D,II= 0.82, E,I= 0, E,II= 0.87
6. Assign examples to the most similar cluster I: (A,B,C) II: (D,E)
7. Repeat steps 3-5 until the clustering got stabilized

Latent Semantic Indexing

- LSI is a statistical technique that attempts to estimate the hidden content structure within documents:
  - …it uses linear algebra technique Singular-Value-Decomposition (SVD)
  - …it discovers statistically most significant co-occurrences of terms

LSI Example

Original document-term matrix

	d1	d2	d3	d4	d5	d6
cosmonaut	1	0	1	0	0	0
astronaut	0	1	0	0	0	0
moon	1	1	0	0	0	0
car	1	0	0	1	1	0
truck	0	0	0	1	0	1

Rescaled document matrix, Reduced into two dimensions

	d1	d2	d3	d4	d5	d6
Dim1	-1.62	-0.60	-0.04	-0.97	-0.71	-0.26
Dim2	-0.46	-0.84	-0.30	1.00	0.35	0.65

High correlation although d2 and d3 don't share any word

Correlation matrix

	d1	d2	d3	d4	d5	d6
d1	1.00					
d2	0.8	1.00				
d3	0.4	0.9	1.00			
d4	0.5	-0.2	-0.6	1.00		
d5	0.7	0.2	-0.3	0.9	1.00	
d6	0.1	-0.5	-0.9	0.9	0.7	1.00

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA

JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

References

References to some of the Books

Reading Material

- M. Grobelnik, D. Mladenić, M. Witbrock. Text Mining for Semantic Web. Encyclopedia of Machine Learning, Sammut and Webb (eds.), Springer-Verlag, 2009.
- M. Grobelnik, D. Mladenić**.** Automated knowledge discovery in advanced knowledge management. Journal of Knowledge management 9:5, 132-149, 2005.
- T.M. Mitchell. Mining Our Reality, Science:326, December 2009.

- M. Grobelnik et al., Machine Learning Techniques for Understanding Context and Process, Context and Semantics for Knowledge Management, 127-148, 2011.
- T.M. Mitchell et al. Populating the Semantic Web by Macro-Reading Internet Text, ISWC-2009.
- M. Grobelnik, D. Mladenić, B. Fortuna. Semantic Technology for Capturing Communication Inside an Organization, *IEEE Internet computing*, 2009, 13:4, 59-66, 2009.

Requirements for this class

- Attendance of the lectures and independent work on the assigned seminar following the provided instructions
- Report on the results of the project work to be sent via e-mail by 15.02.2017 to Blaz.Fortuna@ijs.si
  - 5-10 pages report
- Presentation of the project on 1.03.2017 15:00
  - 5-10 slides presentation (10-15 minutes presentation)
- Oral exam on 1.03.2017